Validity of the Impact Factor of Journals as a Measure of Randomized Controlled Trial Quality

Corrado Barbui, M.D.; Andrea Cipriani, M.D.; Lara Malvini, M.D.; and Michele Tansella, M.D.

Objective: To assess whether the impact factor, a measure of the frequency with which journal articles are cited in the scientific literature, is a proxy measure of the quality of articles reporting the results of randomized controlled trials.

Method: The quality of trials included in an ongoing Cochrane review concerned with the antidepressant fluoxetine was assessed using the Cochrane Collaboration Depression, Anxiety, and Neurosis quality assessment instrument, the Jadad scale, and the quality criterion of the Cochrane Collaboration Handbook. Journal impact factors were extracted from the Journal Citation Report.

Results: A total of 131 articles reported results from 132 clinical trials comparing fluoxetine with other antidepressants. The relationship between trial quality and the impact factor of journals where these studies were published, stratified by period of publication, revealed that journals with impact factors above 4 points published only trials with above-average overall quality ratings, while journals with impact factors below 4 points published both high- and low-quality trials. The Jadad scale revealed similar quality in trials published in journals with high, medium, and low impact factors (Pearson $\chi^2 = 0.298$, $p = .861$), and the quality criterion of the Cochrane Collaboration Handbook showed unclear randomization in the majority of trials and in all 15 trials published in high–impact factor journals (Pearson $\chi^2 = 4.678$, $p = .096$).

Conclusion: The impact factor of journals is not a valid measure of randomized controlled trial quality.

(J Clin Psychiatry 2006;67:37–40)

Received May 13, 2005; accepted July 6, 2005. *From the Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona, Verona, Italy.*

The authors report no financial affiliations or other relationships relevant to the subject of this article.

Corresponding author and reprints: Corrado Barbui, M.D., Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona, Policlinico GB Rossi, 37134 Verona, Italy (e-mail: corrado.barbui@univr.it).

In recent years, the impact factor, a measure of the frequency with which journal articles are cited in the scientific literature, has progressively become the most used indicator of journal quality.[1,2] Despite the limitations of the impact factor as a quality measure,[3–7] clinicians, researchers, and policymakers may assume that high-quality research is published in journals with a high impact factor.

Although there is some evidence that articles of high methodological quality are published in journals with a high impact factor,[8] or in journals given high ratings by clinicians and researchers,[9] so far only general medical journals were analyzed, and research articles adopting different study designs were considered together. In this study, we assessed whether the impact factor is a proxy measure of the quality of randomized controlled trials (RCTs), the reference tool for generating the evidence base that should guide clinicians in reaching a decision about optimal care. An ongoing Cochrane review concerned with the antidepressant fluoxetine included published clinical trials comparing fluoxetine with other antidepressants[10–13] and offered an opportunity for this analysis.

METHOD

Inclusion Criteria

Only studies that randomly allocated patients with major depression to fluoxetine versus any other antidepressant agent were included. Crossover studies and trials in depressed patients with a concomitant medical illness were excluded.

Search Strategy

Randomized controlled trials were located by searching the Cochrane Collaboration Depression, Anxiety, and Neurosis (CCDAN) Controlled Trials Register and the Cochrane Central Register of Controlled Trials. The following search phrase was used: *fluoxetin** **or** *adofen* **or** *docutrix* **or** *erocap* **or** *fluctin* **or** *fluctine* **or** *fluoxeren* **or** *fontex* **or** *ladose* **or** *lorien* **or** *lovan* **or** *mutan* **or** *prozac* **or** *prozyn* **or** *reneuron* **or** *sanzur* **or** *saurat* **or** *zactin*. MEDLINE and EMBASE were searched with no year limits using the search phrase *fluoxetine* **and** *("ran-*

domised controlled trial" **or** *"random allocation"* **or** *"double-blind method")* up to March 2003, which is when the search was performed. Non–English-language articles were included, and reference lists of relevant articles and previous systematic reviews were hand-searched for published reports and citations of unpublished research.

Data Extraction

Two reviewers (A.C., L.M.) independently extracted data; any disagreement was resolved by discussion and consensus with a third member of the team (C.B.). The quality of trials was assessed using the CCDAN quality assessment instrument,[14] the Jadad scale,[15] and the quality criterion of the Cochrane Collaboration Handbook.[16]

The CCDAN instrument, specifically developed for trials of treatments for depression and neurosis, consists of 23 items covering a wide range of aspects of quality including objective formulation, design, presentation of results, analysis, and quality of conclusions. It covers aspects of both internal validity (or control of bias) and external validity (or generalizability). All items equally contribute to the final score. The scale ranges from 0 to 46, with higher scores indicating better quality. The Jadad scale consists of 3 items pertaining to descriptions of randomization, masking, dropouts, and withdrawals in the report of an RCT. The scale ranges from 0 to 5, with higher scores indicating better reporting. The quality criterion of the Cochrane Collaboration Handbook consists of 1 item pertaining to the adequacy of randomization (including its concealment). The scale uses letters to denote quality of randomization (A = adequate, B = unclear, C = inadequate, D = not used).

Journal impact factors from 1984 to 2003 were obtained from the Journal Citation Report of the Institute for Scientific Information.[17] Impact factors were assigned on the basis of the journal and year of publication of each RCT.

Data Analysis

Since the quality of antidepressant trials has progressively improved in the last 4 decades,[18] year of publication was considered a confounder of the relationship between trial quality and impact factor. We therefore stratified RCTs in 4 ways: RCTs published before 1991, RCTs published from 1991 through 1995, RCTs published from 1996 through 2000, and RCTs published from 2001 through March 2003. In each stratum, a Spearman rank correlation coefficient between journal impact factor and CCDAN quality score was calculated. Overall, CCDAN quality score was not categorized because no accepted cutoff values have been described. Medium/high-quality RCTs were defined as those scoring more than 2 out of a maximum of 5 on the Jadad scale, while low-quality RCTs scored 2 or less out of a maximum of 5 on the Jadad scale. These assignments were derived from Moher and colleagues.[19] Impact factors were categorized into the following 3 groups: high impact factor (more than 4 points); medium impact factor (from more than 2 points to 4 points); and low impact factor (from 0 to 2 points). Categorical data were analyzed by Pearson χ^2 statistics.

RESULTS

Trial Characteristics

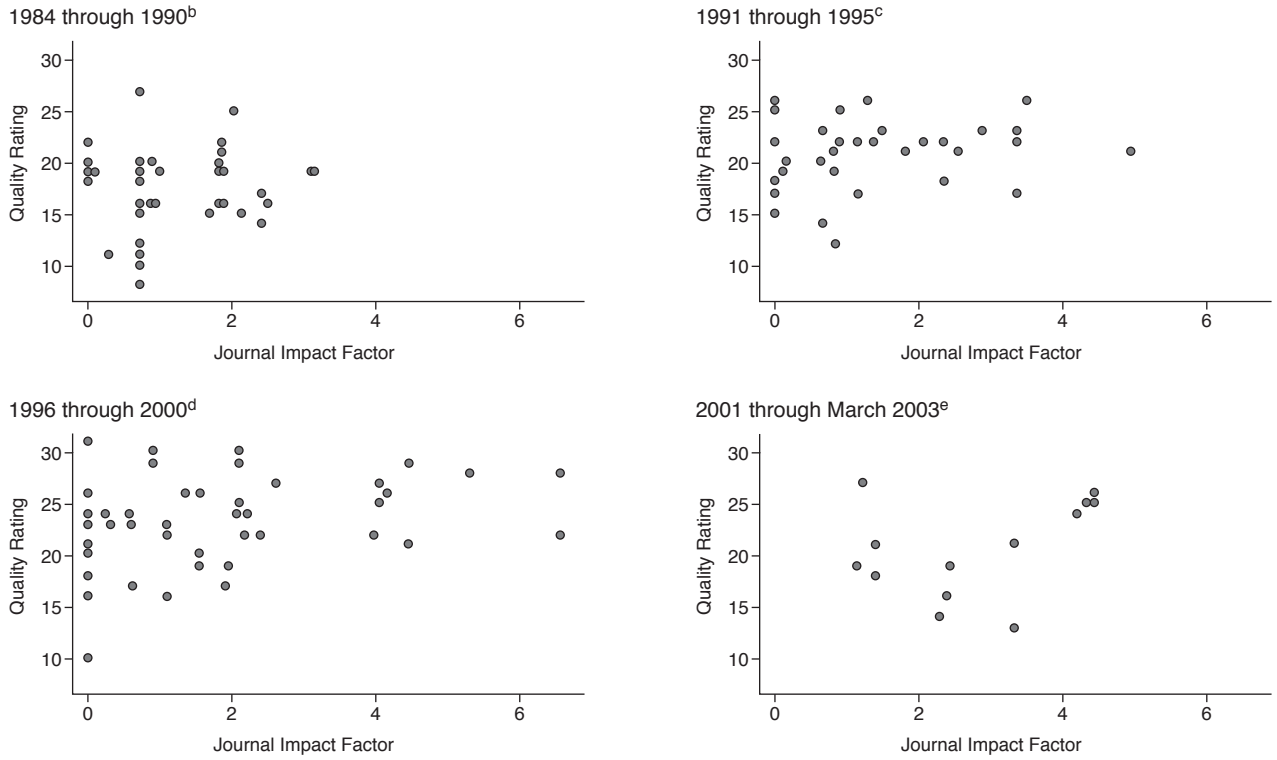
A total of 131 articles reported results from 132 clinical trials comparing fluoxetine with other antidepressants.[10] Most studies were carried out in the United States or in Europe. The mean length of follow-up was 8 weeks (SD = 5.1 weeks). Only 12 trials (9%) were conducted with inpatients, 24 (18%) enrolled both inpatients and outpatients, and the remaining were conducted in outpatient facilities. The majority of studies (72%) enrolled patients meeting DSM-III-R, DSM-IV, or ICD-10 criteria for depression; the others used operational or implicit criteria for depression. Elderly subjects were included in 56 studies (42%). Randomized controlled trials were published in 50 different journals[10]; however, 5 journals accounted for almost 50% of publications. Years of publication ranged from 1984 through March 2003.

Overall CCDAN quality score for the total sample, out of a maximum of 46, ranged from 8 to 31. The mean quality score was 20.7 (SD = 4.67), indicating an overall medium quality of included RCTs. According to the Jadad scale, 95 RCTs (72.0%) were medium/high-quality studies. However, only 8 RCTs scored an "A" according to the quality criterion of the Cochrane Collaboration Handbook.

Relationship Between Trial Quality and the Impact Factor of Journals

The relationship between trial quality and journal impact factor, stratified by period of publication, is presented in Figure 1. In each stratum, journals with impact factors above 4 points published trials with overall quality ratings above the average only, while journals with impact factors below 4 points published both high- and low-quality trials. Only in RCTs published from 1996 through 2000 was the hypothesis that trial quality and journal impact factor were independent rejected (Spearman $\rho = 0.313$, $p = .029$), while in RCTs published in the remaining 3 strata no statistically significant correlation was found. Additionally, the Jadad scale revealed that reporting of randomization, masking, dropouts, and withdrawals was not different between trials published in journals with high, medium, and low impact factors (Pearson $\chi^2 = 0.298$, $p = .861$) (Table 1). The quality criterion of the Cochrane Collaboration Handbook showed unclear randomization in the majority of RCTs and in all 15 trials published in high–impact factor journals (Pearson $\chi^2 = 4.678$, $p = .096$) (Table 1).

Figure 1. Correlation Between Trial Quality and Impact Factor of Journals Where These Trials Were Published, Stratified by Period of Publication[a]



[a]Measured With the Cochrane Collaboration Depression, Anxiety, and Neurosis Scale.
[b]Spearman ρ = 0.014, p = .935.
[c]Spearman ρ = 0.238, p = .174.
[d]Spearman ρ = 0.313, p = .029.
[e]Spearman ρ = 0.430, p = .124.

DISCUSSION

This is the first study investigating the relationship between the impact factor, a measure considered an indicator of journal quality, and the methodological quality of articles reporting clinical trial data. The analysis provided interesting insights that should be interpreted while bearing in mind the different characteristics of the 3 measures of trial quality adopted in this analysis.

The CCDAN quality score has the positive characteristic of covering a wide range of aspects associated with the conduct and reporting of clinical trials, representing, in this way, a suitable instrument for when a general description of quality is warranted.[14] However, this checklist does not adopt any weighting procedure in the calculation of the overall quality score, i.e., all items equally contribute to the final score. Therefore, items investigating the randomization procedure or the concealment of allocation are given the same weight as items investigating side effect reporting or evaluating the reporting of patients' demographic characteristics. The Jadad scale was additionally employed to overcome this limitation. It has the advantage of being focused on key trial characteristics, such as randomization, masking, dropouts, and withdrawals, but obviously does not cover other trial features.[15] Finally, the Cochrane Collaboration Handbook, concerned with randomization only, was employed with the assumption that a detailed description of this procedure represents the core quality characteristic of high-quality clinical trials.[16]

Table 1. Distribution of Trial Quality by the Impact Factor of Journals Where Trials Were Published

	Impact Factor					
	Low (89 RCTs)		Medium (28 RCTs)		High (15 RCTs)	
Quality Measure	No.	%	No.	%	No.	%
Jadad scale						
Low quality	24	27.0	9	32.1	4	26.7
Medium/high quality	65	73.0	19	67.9	11	73.3
Randomization[a]						
Unclear (B)	85	95.5	24	85.7	15	100.0
Adequate (A)	4	4.5	4	14.3	0	

[a]Quality criterion of the Cochrane Collaboration Handbook.

The characteristics of each of these rating instruments help the interpretation of our findings. The analysis

showed that journals with an impact factor above 4 points published clinical trials with an overall quality rating above the average. However, clinicians, researchers, and policymakers should be aware that the converse was not true for journals with impact factors below 4 points: the quality of clinical trials published in these journals varied from low to high. In addition, when a strict definition of quality was adopted, no trend was noticed. Likewise, when randomization, a key feature for comparing the outcomes of treatments given to groups of patients that do not differ in any systematic way,[20] was the only quality criterion, none of 15 RCTs published in high–impact factor journals provided enough detail to receive an "A" rating. We therefore conclude that the impact factor of journals is not a valid measure of RCT quality. Articles reporting clinical trial data in high–impact factor journals might better report some ancillary information on study design and trial characteristics, but they fail to better report key details on randomization and its concealment.

It would be of interest to replicate this analysis in other psychiatric conditions, such as schizophrenia, and in other homogeneous samples of trials investigating nonpharmacologic interventions, in order to clarify whether these results apply to the reporting of clinical trials in general or whether they specifically apply to antidepressant RCTs.

Drug name: fluoxetine (Prozac and others).

REFERENCES

1. Leff D. Making an impact: the rise of the impact factor as a measure of journal quality. J Am Diet Assoc 2005;105:29–30
2. Garfield E. Journal impact factor: a brief review [editorial]. CMAJ 1999;161:979–980
3. Seglen PO. Why the impact factor of journals should not be used for evaluating research. BMJ 1997;314:498–502
4. Bloch S, Walter G. The impact factor: time for change. Aust N Z J Psychiatry 2001;35:563–568
5. Moed HF, van Leeuwen TN. Impact factors can mislead [letter]. Nature 1996;381:186
6. Porta M. Quality of impact factors of general medical journals: quality matters—and the choice of indicator matters too [letter]. BMJ 2003;326:931
7. Amin M, Mabe MA. Impact factors: use and abuse. Medicina (B Aires) 2003;63:347–354
8. Lee KP, Schotland M, Bacchetti P, et al. Association of journal quality indicators with methodological quality of clinical research articles. JAMA 2002;287:2805–2808
9. Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? J Med Libr Assoc 2003;91:42–46
10. Patten S, Cipriani A, Brambilla P, et al. International dosage differences in fluoxetine clinical trials. Can J Psychiatry 2005;50:31–38. Available at: http://www.cpa-apc.org/Publications/Archives/CJP/2005/January/cjp-jan05-barbui-RP.pdf. Accessibility verified Dec 13, 2005
11. Barbui C, Cipriani A, Brambilla P, et al. "Wish bias" in antidepressant drug trials? J Clin Psychopharmacol 2004;24:126–130
12. Brambilla P, Cipriani A, Hotopf M, et al. Side-effect profile of fluoxetine in comparison with other SSRIs, tricyclic and newer antidepressants: a meta-analysis of clinical trial data. Pharmacopsychiatry 2005;38:69–77
13. Furukawa TA, Cipriani A, Barbui C, et al. Imputing response rates from means and standard deviations in meta-analyses. Int Clin Psychopharmacol 2005;20:49–52
14. Moncrieff J, Churchill R, Drummond DC, et al. Development of a quality assessment instrument for trials of treatments for depression and neurosis. Int J Methods Psychiatr Res 2001;10:126–133
15. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1–12
16. Sackett D. The Cochrane Collaboration Handbook. Oxford: Update Software; 1997
17. Science Citation Index: Journal Citation Report 1984–2003. Philadelphia, Pa: Institute for Scientific Information
18. Barbui C, Hotopf M. Forty years of antidepressant drug trials. Acta Psychiatr Scand 2001;104:92–95
19. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609–613
20. Altman DG, Bland JM. Statistics notes: treatment allocation in controlled trials: why randomise? BMJ 1999;318:1209